Gkerekos, Christos and Lazakis, Iraklis and Theotokatos, Gerasimos (2016) Ship machinery condition monitoring using vibration data through supervised learning. In: Proceedings of MSO 2016, International Conference on Maritime Safety and Operations. University of Strathclyde Publishing, pp. 103-110. ISBN 9781909522169 ,

Ship Machinery Condition Monitoring using Vibration Data through Supervised Learning

C. Gkerekos, I. Lazakis & G. Theotokatos
Department of Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, Glasgow, UK

ABSTRACT: This paper aims to present an integrated methodology for the monitoring of marine machinery using vibration data. Monitoring of machinery is a crucial aspect of maintenance optimisation that is required for the vessel operation to remain sustainable and profitable. The proposed methodology will train models using pre-classified (healthy/faulty) data and then classify new data points using the models developed. For this, vibration points are first acquired, appropriately processed and stored in a database. Specific features are then extracted from the data and stored. These data are then used to train supervised models pertinent to specific machinery components. Finally, new data are compared against the models developed in order to evaluate their condition. The above will provide a flexible but robust framework for the early detection of emerging machinery faults. This will lead to minimisation of ship downtime and increase of the ship's operability and income through operational enhancement.

KEYWORDS: vibration measurements; predictive maintenance; machine learning; condition monitoring; SVM

1 INTRODUCTION

Over four-fifths of the world merchandise trade are carried by sea (United Nations, 2015). This highlights the significance of ships in the global goods transportation system. Meanwhile, the current financial situation of the shipping industry, certainly not aided by the drop in oil prices, has led to a drop in chartering rates. More so, the global merchant fleet has an average age of almost 20 years (United Nations, 2015). With profit margins shrinking and average vessel age rising, it is clear that a high level of optimisation is required for the vessel operation to remain sustainable and profitable.

Optimisation of machinery components' maintenance can majorly contribute to the sustainability and profitability of a vessel. Specifically, this optimisation concerns the selection of aspects from both preventive and reactive (run-to-failure) maintenance for uptime maximisation and cost minimisation. In the former maintenance type, availability and reliability are kept high; albeit compromising with high part-replacement costs and increased (planned) downtime due to relevant replacement work. In the latter maintenance type, the profit obtained through the decreased planned maintenance actions is counterbalanced by the high cost – both in time and in financial sense – of replacement after a component fails completely. An optimised maintenance scheme, usually called Predictive Maintenance (PdM), aims to combine the best aspects of preventive and reactive in order to suggest maintenance actions at the right time and offer the greatest gains in terms of cost and time.

Predictive maintenance requires some form of input in order to be able to deduce whether maintenance is at that time required and, if not, forecast when it will be. Past research (Wang et al., 2015, Wang et al., 2016b, Rashid et al., 2016) has shown that vibration measurements from machinery components can accurately convey the condition of a component after suitable analysis of the acquired signals.

While Predictive maintenance is widely used in other fields such as nuclear power production and aerospace, there are not many applications in the marine field. In the marine field, most maintenance actions are performed based on Original Equipment Manufacturers' (OEM) recommendations, i.e. a form of Preventive maintenance.

Hence, this paper aims to present the development of an integrated framework concerning firstly the storage and suitable processing of vibration data and secondly the training of appropriate models for the condition monitoring of marine machinery.

Section 1 introduces the paper's scope and motivation of research. Section 2 refers to the research background, including a summary of maintenance types and an overview of past research on machinery condition monitoring. Section 3 elaborates on the proposed methodology concerning the data processing and storage as well as model training. Section 4 details the setup of two case studies used to validate the proposed methodology. Section 5 presents and discusses the results obtained through these case studies. Finally, in section 6, overall conclusions are provided along with further research steps.

2 RESEARCH BACKGROUND

In general, three types of maintenance are applicable for machinery applications: reactive, preventive, and predictive or condition-based.

Reactive maintenance concerns maintenance that is only performed following the complete failure of a component. At that point, no repairing is possible and the component is replaced by a new one (Mohanty, 2015). This method of maintenance provides the longest time between shutdowns but failures are catastrophic and can possibly affect multiple components and/or machines (Randall, 2011). Hence, reactive maintenance is mainly applied to relatively not expensive and non-critical machines or where redundancies exist.

Preventive maintenance refers to maintenance that happens at a fixed frequency, usually following Original Equipment Manufacturer (OEM) recommendations. As preventive maintenance generally aims to provide maintenance intervals such that no more than a 1-2% probability of experiencing machinery failures, the vast majority of machines would be able to continue working without maintenance for two or three intervals (Neale & Woodley, 1978). This has been shown to cause reduced morale in maintenance workers and introduce increased "infant mortality" in machines due to faults than would otherwise have been avoided (Randall, 2011).

Predictive maintenance provides a more intelligent method of maintenance planning. There, present and past condition of each component is taken into consideration in order to offer bespoke maintenance scheduling for each component and each machine. Predictive maintenance requires a higher expenditure at installation but over an extended period of time, becomes more economical that preventive or reactive maintenance. Especially in industries where machines are expected to run for long periods without any shutdowns, it was shown that predictive maintenance can reduce relevant costs by up to 65% (Neale & Woodley, 1978). Jardine et al. (2006) provide an overview of diagnostics and prognostics methods applicable to predictive maintenance.

In order to estimate past and present condition of components and machines, different types of measurements are acquired and processed. Depending on the application and the data available, the main types of input data considered are the following: vibration measurements; acoustic measurements; oil/debris analysis; corrosion (thickness) measurements; thermography; motor current signature analysis; and performance monitoring (Scheffer & Girdhar, 2004, Mohanty, 2015).

2.1 Maintenance in the maritime sector

In sectors such as defence, aviation, manufacturing, automobile and nuclear power production, maintenance focus has recently shifted from reactive to preventive/predictive (Lazakis et al., 2010).

Ship maintenance amounts to 10-15% of the shipping company direct operating costs (Stopford, 2008). However, in the maritime sector, ship maintenance has been considered an area of needless expenditure and advanced monitoring methods have not yet been applied to a large extent (Lazakis & Olcer, 2015).

Nevertheless, some attempts towards predictive maintenance in shipping have been made in the past few years. For example, Chandroth (2004) proposed a system where vibration data are combined with cylinder pressures for the condition monitoring of a main engine. Accordingly, Lus (2013) suggested a methodology for the monitoring of high-speed diesel engines using vibration data from valve gear mechanisms and fuel systems. Furthermore, guidelines on vibration monitoring have been proposed by relevant maritime bodies (Reed et al., 1987, Det Norske Veritas, 2011).

2.2 Vibration monitoring of machinery

Vibration-based condition monitoring is a well-studied topic with multiple research publications. Randall (2011) provides an overview of vibration monitoring methods including signal processing and diagnostics methods. Ruiz-Carcel et al. (2016) proposed a method of merging process and vibration data in order to detect mechanical faults in machinery working under variable conditions. Al-Badour et al. (2011) proposed an analogous method of combining vibration measurements from multiple sensors for the condition monitoring of electric motors. Egusquiza et al. (2015) described a method of combining acceleration and pressure measurements with operational data for the condition monitoring of pump-turbines. Yin & Hou (2016) and Widodo & Yang (2007) both presented an overview of Support Vector Machine (SVM) techniques for fault diagnosis and monitoring in engineering applications.

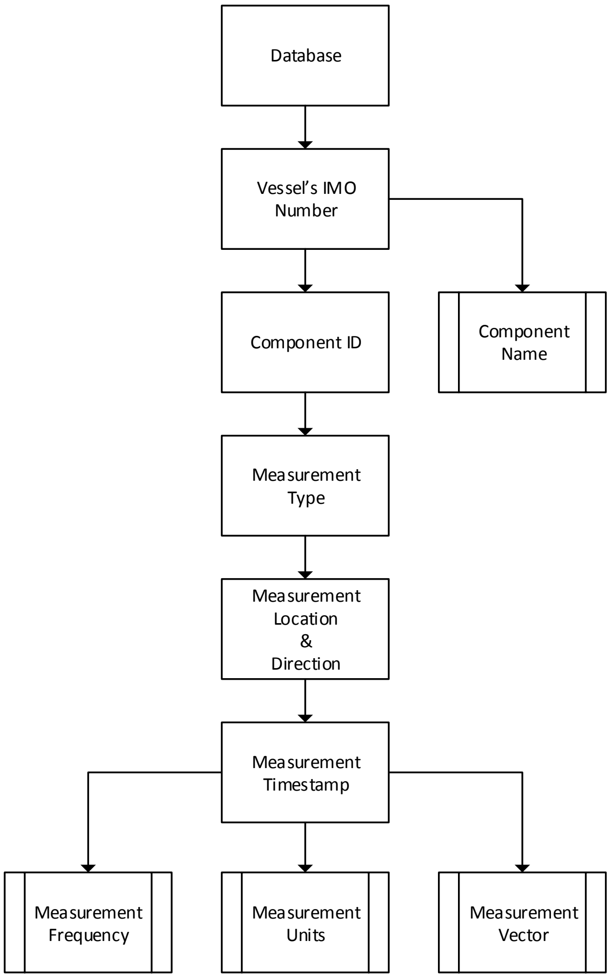

Figure 1. Database visual representation.

3 SUGGESTED METHODOLOGY

The methodology elaborated in this paper concerns a) the design of a database for suitable storage of measurements obtained from ship machinery for condition monitoring purposes and b) the use of stored measurements for the detection of incipient faults and anomalies.

3.1 Data storage

For the design of the database, the nature of each measurement (vector vs. scalar) as well as any particular data access needs are considered. Based on these, a database based on the schematic shown in Figure 1 was designed.

At top-level, data are classified by vessel's IMO number. As each component has a unique ID that is included in each measurement file, machinery components are then classified by ID. At the same time, the component name is stored in a string to simplify data retrieval and lookup by name. As this database was designed to accommodate any kind of measurement relating to machinery condition monitoring, the next level stores the measurement type. Possible types include temperature, ultrasonic and vibration measurements. Depending on the properties of each component, vibration measurements are either recorded in velocity or acceleration units. Next, the location of each measurement is noted. Measurements in different locations and directions are considered for each component/machine. For example, in the case of a main engine, as depicted in Figure 2, measuring points fore and aft of the engine, as well as port and starboard are considered. Accordingly, in the case of diesel generators or pumps, measurements at both free ends as well as the shaft coupling are considered. Additionally, in the case of vibration measurements, at each location measurements in various directions (i.e. vertical, horizontal and/or axial) are acquired. For each location and direction, the timestamp of each measurement is stored along with the sampling frequency units (important especially in the case of acceleration measurements where both m/s^2 and "g" units are acceptable) and a vector that contains the actual measurements obtained.

The proposed hierarchical database structure simplifies data access for both plotting and model training.

3.2 Data pre-processing

Acquired data need to be pre-processed before being used for model training. Pre-processing is currently a hot topic in data mining and predictive analytics, with cutting edge research focusing on optimised pre-processing algorithms (Garcia et al., 2015).

As an alternative to advanced algorithms, some visual red flags for determining whether a dataset requires pre-processing have also been proposed, e.g. as shown in Table 1.

If any measurement data points are originally missing or eventually get discarded due to a red flag, these points can be accordingly substituted. A number of methods for treating these data points exist (Kotsiantis et al., 2006, Lakshminarayan et al., 1996). A straight-forward approach is to completely discard any instance that contains a missing feature. Alternatively, any missing feature can be replaced by the mean or mode (i.e. most commonly found) value of that feature, taking into consideration the whole dataset. Another proposed method concerns the training of a regression/classification model using the remaining dataset points as training input and using all known instance features as model input so that the unknown feature value is obtained as model output. Finally, the "hot deck imputation" method concerns the identification of the most similar instance and substituting the unknown feature value with that of the identified instance.

If the acquired dataset is imbalanced and there exist over-represented or under-represented classes, either under-represented data points can be duplicated or over-represented data points can be removed (Kotsiantis et al., 2006).

Table 1. Examples of data pre-processing red flags (Kotsiantis et al., 2006).

Problems	Metadata	Examples/Heuristics
Illegal values	Cardinality	e.g., cardinality (gender) > 2 indicates problem
	Max, min	Max, min should not be outside permissible range
	Variance, deviation	Variance, deviation of statistical values should not be higher than threshold
Misspellings	Feature values	Sorting on values often brings misspelled values next to correct values

The suggested methodology concerning the pre-processing of the acquired and stored data follows Table 1.

As all recorded values either contain measured numerical values of descriptions in short-text form, there is no expectation of cardinality issues.

Range issues can sometimes be expected in machinery condition monitoring applications, especially when sensors malfunction or function outside their operational range. Depending on the context of each measurement, determining an expected measurement range and discarding any outlying values is possible.

Another way of determining whether a measurement should be discarded is by observing specific values along with the general trend observed and comparing with the expected process speed. Accordingly, the variance of each feature was calculated and compared to a pre-decided threshold.

Finally, in the case of measurement location and direction fields, a closest-match algorithm was designed to mitigate against misspellings. Using this algorithm, instead of creating a new location record for each misspelling, the data is stored in the closest match. This is based on the fact that measuring points are predefined in the sense that the location is defined by a number as shown in Figure **2** and the direction by the letters x, y or z. As such, the algorithm can easily map, e.g. "P3 – z" to "3z".

3.3 Feature extraction

Vibration measurements are high-frequency time-domain measurements acquired either constantly or intermittently at intervals that depend on the criticality and specific properties of each component (Randall, 2011). While in some cases measurement intervals can be obtained empirically, novel model-based approaches have also been proposed (Sherwin & Al-Najjar, 1999).

An inherent difficulty of vibration monitoring is the need to map a significantly large time-series (e.g. a 10-second recording at 100 kHz sampling rate corresponds to 1 million points) into some quantity than can be easily managed and stored. At the same time, this quantity needs to be a metric that accurately conveys the condition of a component.

Multiple such metrics have been proposed, with the most pertinent to machinery condition monitoring assembled in Table 2 (Kharce & Kshirsagar, 2014, Lamraoui et al., 2015, Pascual, 2015, Wang et al., 2015). These metrics (features) are divided firstly into those that take as input the time-domain signal and those that first require a transformation to frequency-domain. In the notation followed in Table 2, x corresponds to the time-series measurement, with x_i referring to the i-th element of x. Accordingly, f_i and p_i respectively correspond to the i-th frequency and amplitude of the frequency-domain transformation.

3.4 Dimensionality reduction

In general, the number of available training samples (N) should exceed the number of features (L) as the complexity of a model cannot exceed the complexity of the training dataset. Depending on the application, the features' nature and specific assumptions, different N/L ratio recommendations are provided in literature (Foley, 1972, Hua et al., 2005).

Often, the derived features are not linearly uncorrelated, i.e. one depends on another or both depend on the same variables. In such cases, dimensionality reduction techniques, such as Principal Component Analysis (PCA) are applied (Skittides & Früh, 2014, Wang et al., 2016a).

PCA provides the orthogonal transformation of possibly correlated variables into a set of linearly uncorrelated ones (principal components). To perform PCA on a given dataset, the following steps are required (Maimon & Rokach, 2010, Smith, 2002):

1. The mean of each data dimension is calculated and subtracted from the original dataset in order to obtain the adjusted dataset.
2. The dataset's covariance matrix C is calculated.
3. Eigenvalues and unit eigenvectors of the covariance matrix C are calculated.
4. Eigenvectors are sorted by eigenvalue, highest to lowest. A number n of features is selected based on the explained variance – complexity trade-off and a feature vector is obtained by combining the first n eigenvectors.
5. The post-PCA dataset matrix is obtained by multiplying the transpose of the feature vector by the transpose of the adjusted dataset matrix.

Table 2. Time-domain and frequency-domain features extractable from vibration datasets (Kharce & Kshirsagar, 2014, Lamraoui et al., 2015, Pascual, 2015, Wang et al., 2015).

		Feature	Definition
Time-domain	Dimensional	Root mean square	$RMS = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$
		Mean value	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
		Root value	$Rv = \left(\frac{1}{n}\sum_{i=1}^{n} \lvert x_i\rvert^{\frac{1}{2}}\right)^2$
		Standard deviation	$Std = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{\frac{1}{2}}$
		Kurtosis value	$Kuv = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{(n-1)\times Std^4}$
		Skewness value	$Skv = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(n-1)\times Std^3}$
		Peak value	$Pv = \frac{1}{2}\,(\max(x)-\min(x))$
	Dimensionless	Crest factor	$Cf = \frac{Pv}{RMS}$
		Shape factor	$Shf = \frac{RMS}{\bar{x}}$
		Impulse factor	$Imf = \frac{Pv}{\bar{x}}$
		Clearance factor	$Clf = \frac{Pv}{Rv}$
		Skewness factor	$Skf = \frac{Skv}{RMS^3}$
		Kurtosis factor	$Kuv = \frac{Kuv}{RMS^4}$
Frequency-domain		Mean frequency	$Mf = \sum_{i=1}^{n}\frac{p_i}{N}$
		Frequency centre	$Fc = \frac{\sum_{i=1}^{n} f_i\, p_i}{\sum_{i=1}^{n} p_i}$
		Root mean square frequency	$RMSf = \left(\frac{\sum_{i=1}^{n} f_i^2\, p_i}{\sum_{i=1}^{n} p_i}\right)^{\frac{1}{2}}$
		Root variance frequency	$Rvf = \left(\frac{\sum_{i=1}^{n}(f_i-Mf)^2\, p_i}{\sum_{i=1}^{n} p_i}\right)^{\frac{1}{2}}$

By only retaining a number of features, the dimensionality of the modelling is reduced providing a basis for better results given a limited amount of samples n. Nevertheless, at the same time there exists the inherent trade- off of some variance from the dataset being lost. As such, an optimal number of principal components should be selected so that most of the dataset's variance remains explained while the total number of features is reduced.

3.5 Classifier training

Once a number of principal components are derived through PCA, these are used to train a classifier model that is expected to be able to discern between two classes of data: healthy and faulty. Neural Networks (NN), Support Vector Machines (SVM) and Decision trees are the most common algorithms used for classification based on machine learning.

One of the most important advantages of SVM compared to other pertinent algorithms is its suitability when treating small datasets (Yin & Hou, 2016). SVMs are based on Structural Risk Minimisation (SRM) that leads to balancing model complexity against overfitting (Vapnik, 1989). Another advantage is that any minimum achieved through SVM will be a global minimum, something that is not necessarily true when treating NN minima.

SVMs aim to create a separating hyperplane that is able to classify input data as either positive (healthy in the case of CM) or negative (i.e. faulty). This can either be done linearly or through a non-linear kernel function. In the case of non-linear kernel functions, input data are mapped to a high-dimensional feature-space where linear classification is possible.

The exact mathematical formulation of SVM is not elaborated in this paper but the reader can refer to either Vapnik's own work (Vapnik, 1989) or any book in Data Mining, e.g. Maimon & Rokach (2010).

4 METHODOLOGY APPLICATION

In this section, two case studies are presented. The first concerns database design and data storage. For this, vibration measurements from ship machinery acquired in the framework of EU FP7 INCASS (Inspection Capabilities for Enhanced Ship Safety) project measuring campaign are parsed and stored. The second case study treats the subject of anomaly detection through the training of a suitable classifier. For this, wind turbine gear vibration measurements are used. Both case studies were tested in MATLAB environment.

4.1 Case study I (Data storage)

This case study involves the parsing and storing of vibration measurements acquired on board a tanker vessel and a containership as part of INCASS project measuring campaigns (Raptodimos et al., 2016). Table 3 contains the description of the machinery considered in this case study. For each machinery system, measurements in multiple directions were acquired. As an example, the measuring points for a main engine are shown in Figure 2.

Measurement datasets are contained in Machinery Condition Monitoring (MCM) files. MCM files are based on the xml structure and were developed by the INCASS consortium to provide a streamlined method of recording and exchanging machinery condition monitoring datasets (Taheri et al., 2015).

In order to access and store the datasets contained in MCM files, an .xml parser fine-tuned to the requirements of MCM format was designed.

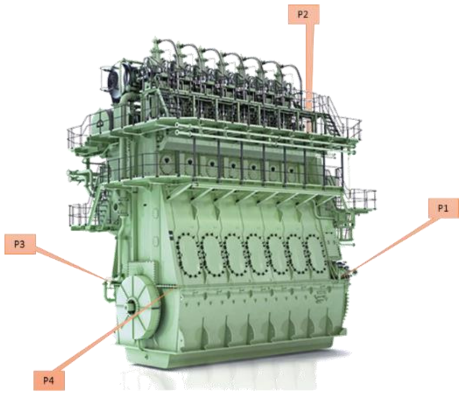

Figure 2. Main engine measuring points.

Table 3. Ship machinery considered in case study 1.

Machine description	Tanker Vessel	Containership
S.W. pump	x	x
C.F.W. pump	x	x
Diesel generator	x	x
Air compressor	x	x
Main engine	x	x
Air blowers	x	x
F.D. fan	x	
F.O. purifier	x	
L.O. purifier	x	
L.O. boost pump	x	
Bilge pump		x

4.2 Case study II (anomaly detection)

The second case study involves the validation of the anomaly detection methodology using radial vibration measurements from a 3 MW wind turbine pinion gear, as suitable data from a ship machinery application were not available. Two datasets were provided, all obtained from wind turbines of the same type and model. One dataset included vibrations on 11 occasions from a pinion that proved to be faulty and the other dataset measurements on 13 occasions from pinions that were assumed as healthy. This dataset is suitable for the evaluation of the classifier methodology proposed above.

The provided datasets were both observed to be clean (i.e. without corrupt or inaccurate points), thus no pre-processing was required. Features were derived as per Table 2 and then PCA was performed to reduce the number of features required for training. SVMs with different kernel methods were trained and compared.

As only a limited amount of data was available, splitting the dataset into training, validation and testing sub-datasets was impractical. Instead a cross validation scheme was applied to evaluate the soundness of the methodology. In cases of limited datasets, the leave-one-out cross validation technique is generally proposed (Wong, 2015). Leave-one-out cross validation is a special case of k-fold validation where the number of folds equals the number of available samples (i.e. the number of measurement occasions in this case).

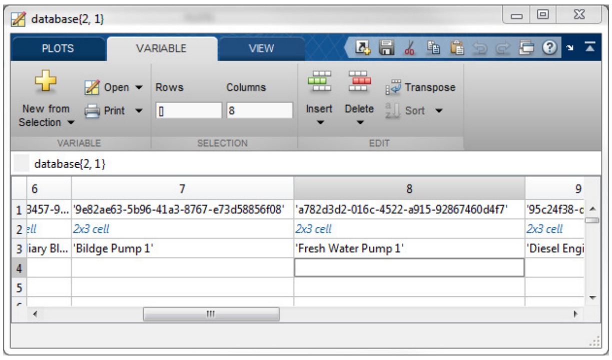

Figure 3. Database snapshot, showing component names (row 3) and IDs (row 1).

5 RESULTS

This section presents the results obtained through the two case studies described above. For the data storage case study, snapshots of the database are provided. Additionally, some results concerning the time requirements of MCM parsing along with compression capabilities of the database are shared. For the anomaly detection case study, results from the PCA are presented, along with results obtained through the trained classifier.

5.1 Case study I (Data storage)

Parsing a full set of machinery measurements on board a vessel lasted approximately 300 seconds. At the same time, measurements with a file-size of almost 300 MB were stored in a 50 MB database.

In Figures 3 and 4, snapshots from the database are shown. In Figure 3, recordings are presented at a component level, with each column corresponding to a different component and each row corresponding to different bits of information. Row 1 contains the component ID and row 3 the relevant component description/name. Row 2 leads to the next database level, i.e. measurement type. In Figure 4, a lower level is presented. There, a specific component ID, measurement type and measurement location/direction have been selected. Hence, at this point, each column refers to a different time-instance. Row 1 contains the time-stamp of the recording, row 3 contains the sampling rate used and row 4 corresponds to the units of the measured signal.

This case study showcased an optimised data storage methodology, obtaining an 85% decrease in file size. At the same time, measurements are easily accessible for model training and plotting.

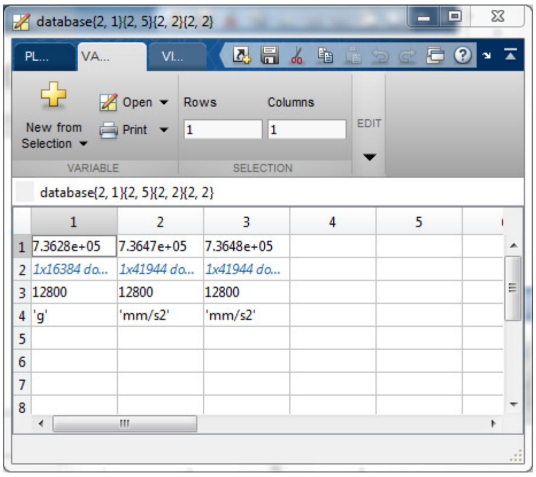

Figure 4. Database snapshot, showing measurement timestamps (row 1), frequency (row 3) and units (row 4). Measurement vector in row 2.

5.2 Case study II (anomaly detection)

Performing PCA on the acquired dataset, it was observed that several features were highly correlated and that a limited amount of principal components sufficed to accurately describe the initial dataset. This can be observed in the Pareto chart shown in Figure 5. A single principal component explained over 70% of the dataset variance, whilst considering three features, over 95% of the original variance was explained. All SVM where trained using the first three principal components.

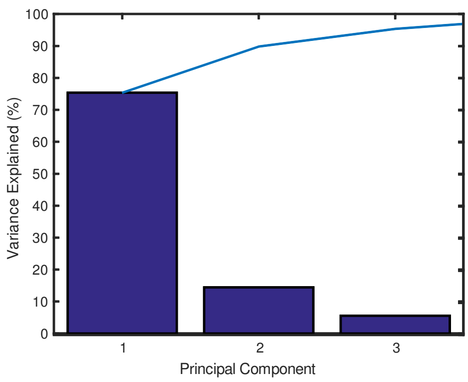

Figure 5. Variance explained vs. number of principal components. Bars show the variance explained by the i-th component while the solid line demonstrates cumulative variance.

Table 4. Results obtained through the evaluation of different SVM kernels.

SVM Kernel	Accuracy attained
Linear	58.3%
Quadratic	91.7%
Cubic	95.8%

Following this, SVMs based on different kernels were trained and evaluated. The results obtained are presented in Table 4. Cubic SVM presented the best results overall, achieving a 95.8% accuracy.

The results obtained through the trained model are also depicted in Figure 6 in the form of a scatter plot.

This case study demonstrated that accurate models for machinery fault detection can be trained using support vector machines. Of special note is the fact that no prior knowledge of the machinery examined was required as input apart from vibration measurements.

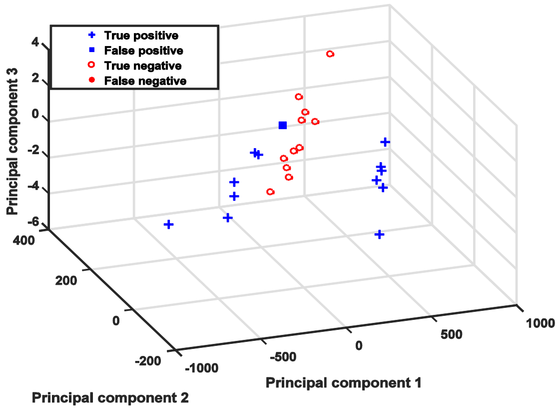

Figure 6. Performance of Cubic SVM classifier.

6 CONCLUSIONS

This paper aims to present a framework for the storage, pre-processing and suitable processing of vibration measurements for anomaly detection in ship machinery. First, an overview of the current state of research in the field of maritime maintenance and condition monitoring was provided. Then, proposed methodology was elaborated and validated through two case studies. These case studies respectively concerned data storage and model training for fault detection.

In conclusion, future research steps include further development of the anomaly detector/classifier so that the discernment between different faults is possible. In addition, the development and implementation of a Decision Support System providing guidance with regards to the selection of optimal maintenance actions is proposed. Moreover, the methodology will be further validated by repeating case study II using suitable data from ship machinery. Additionally, the use of performance alongside vibration measurements will be considered to allow for accurate monitoring under variable operating conditions.

ACKNOWLEDGEMENTS

The work presented in this paper is partially funded by INCASS project. INCASS Project has received research funding from the European Union's Seventh Framework Programme under grant agreement no 605200. This publication reflects only the author’s views and European Union is not liable for any use that may be made of the information contained herein.

Acknowledgement is also made for the wind turbine gear vibration measurements used in this work provided through data-acoustics.com Database.

REFERENCES

Al-Badour, F., Sunar, M. & Cheded, L. 2011. Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. Mechanical Systems and Signal Processing, 25**,** 2083-2101.

Chandroth, G. 2004. Condition monitoring: the case for integrating data from independent sources. Journal of Marine Engineering & Technology, 3**,** 9-16.

Det Norske Veritas 2011. Vibration Class. Rules for Classification of Ships - Newbuildings - Special equipment and Systems - Additional class (Part 6, Chapter 15).

Egusquiza, E., Valero, C., Valentin, D., Presas, A. & Rodriguez, C. G. 2015. Condition monitoring of pump-turbines. New challenges. Measurement, 67**,** 151-163.

Foley, D. H. 1972. Consideration of Sample and Feature Size. IEEE Transactions on Information Theory, IT-18.

Garcia, S., Luengo, J. & Herrera, F. 2015. Data Preprocessing in Data Mining, Springer.

Hua, J., Xiong, Z., Lowey, J., Suh, E. & Dougherty, E. R. 2005. Optimal number of features as a function of sample size for various classification rules. Bioinformatics, 21**,** 1509-15.

Jardine, A. K. S., Lin, D. M. & Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20**,** 1483-1510.

Kharce, P. P. & Kshirsagar, S. V. 2014. Review of Fault Detection in Rolling Element Bearing. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1**,** 169-174.

Kotsiantis, S. B., Kanellopoulos, D. & Pintelas, P. E. 2006. Data Preprocessing for Supervised Learning. International Journal of Computer Science, 1**,** 111-117.

Lakshminarayan, K., Harp, S. A., Goldman, R. & Samad, T. 1996. Imputation of missing data using machine learning techniques. The Second International Conference on Knowledge Discovery and Data Mining. Portland, Oregon.

Lamraoui, M., El Badaoui, M. & Guillet, F. 2015. Chatter Detection in CNC Milling Processes Based on Wiener-SVM Approach and Using Only Motor Current Signals. Vibration Engineering and Technology of Machinery, Proceedings of VETOMAC X 2014, held at the University of Manchester, UK, September 9-11, 2014. Springer International Publishing.

Lazakis, I. & Olcer, A. 2015. Selection of the best maintenance approach in the maritime industry under fuzzy multiple attributive group decision-making environment. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment.

Lazakis, I., Turan, O. & Aksu, S. 2010. Increasing ship operational reliability through the implementation of a holistic maintenance management strategy. Ships and Offshore Structures, 5**,** 337-357.

Lus, T. 2013. Tests results of new diagnostic method for WOLA marine diesel engines type H. Journal of KONES, 20.

Maimon, O. & Rokach, L. 2010. Data Mining and Knowledge Discovery Handbook, Springer.

Mohanty, A. R. 2015. Machinery Condition Monitoring: Principles and Practices, CRC Press.

Neale, M. J. & Woodley, B. J. 1978. A guide to the condition monitoring of machinery - Report TRD 223 for the British Department of Industry

Pascual, D. G. 2015. Artificial Intelligence Tools: Decision Support Systems in Condition Monitoring and DIagnosis, CRC Press.

Randall, R. B. 2011. Vibration-based Condition Monitoring, John Wiley & Sons.

Raptodimos, Y., Lazakis, I., Theotokatos, G., Salinas, R. & Moreno, A. 2016. Collection & Analysis of Data for Ship Condition Monitoring aiming at Enhanced Reliability & Safety. Ocean and Polar Engineering (ISOPE), The 26th International Conference on. Rhodes, Greece.

Rashid, M. M., Amar, M., Gondal, I. & Kamruzzaman, J. 2016. A data mining approach for machine fault diagnosis based on associated frequency patterns. Applied Intelligence.

Reed, E. F., Antonides, G. P. & Lundgaard, B. 1987. Guidelines for the Use of Vibration Monitoring for Preventive Maintenance. In: VIBRATIONS), P. M.-M. (ed.). Society of Naval Architects and Marine Engineers.

Ruiz-Carcel, C., Jaramillo, V. H., Mba, D., Ottewill, J. R. & Cao, Y. 2016. Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions. Mechanical Systems and Signal Processing, 66-67**,** 699-714.

Scheffer, C. & Girdhar, P. 2004. Practical Machinery Vibration Analysis and Predictive Maintenance, Newnes.

Sherwin, D. J. & Al-Najjar, B. 1999. Practical models for condition monitoring inspection intervals. Journal of Quality in Maintenance Engineering, 5**,** 203-221.

Skittides, C. & Früh, W.-G. 2014. Wind forecasting using Principal Component Analysis. Renewable Energy, 69**,** 365-374.

Smith, L. I. 2002. A tutorial on Principal Components Analysis. http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

Stopford, M. 2008. Maritime Economics, Routledge.

Taheri, A., Lazakis, I. & Koch, T. Catalogue Data Model Based Innovative Database System for Ship Structural and Machinery Condition Monitoring (MCM) Data Exchange Protocol. Shipping in Changing Climates (SCC), International Conference on 2015 Glasgow, UK.

United Nations 2015. Review of Maritime Transport. United Nations Conference on Trade and Development.

Vapnik, V. N. 1989. Statistical Learning Theory, Wiley-Interscience.

Wang, X., Zheng, Y., Zhao, Z. & Wang, J. 2015. Bearing Fault Diagnosis Based on Statistical Locally Linear Embedding. Sensors (Basel), 15**,** 16225-47.

Wang, Y., Ma, X. & Joyce, M. J. 2016a. Reducing sensor complexity for monitoring wind turbine performance using principal component analysis. Renewable Energy, 97**,** 444-456.

Wang, Y. X., Xiang, J. W., Markert, R. & Liang, M. 2016b. Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. Mechanical Systems and Signal Processing, 66-67, 679-698.

Widodo, A. & Yang, B. S. 2007. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21**,** 2560-2574.

Wong, T.-T. 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition, 48**,** 2839-2846.

Yin, Z. & Hou, J. 2016. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. Neurocomputing, 174**,** 643-650.